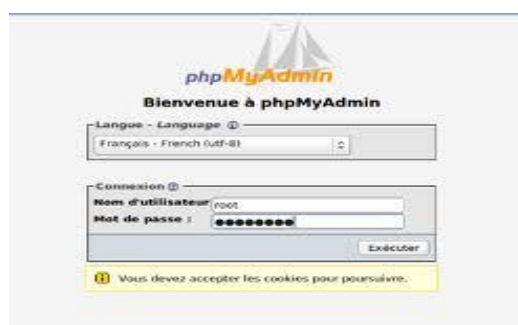


IUT CLERMONT
Antenne d'AURILLAC
GENIE BIOLOGIQUE OPTION
BIOINFORMATIQUE



RAPPORT DE STAGE DE FIN D'ETUDES

CONCEPTION D'UNE BASE RECUEILLANT DES DONNEES AGRONOMIQUES



Maître de stage : Christophe POSER,
Agronome CIRAD
Tuteur : Jocelyne HURST,
Professeur de communication

PALAMARA Frédéric
Stage effectué à Saint Denis de La
Réunion

Du 04/04/2011 au 22/08/2011

UR Systèmes de culture annuels

Sommaire

REMERCIEMENTS

RESUME

ABSTRACT

INTRODUCTION.....1

1. Contexte..... 3

1.1 Le Cirad et l'île de La Réunion 3

1.1.1 Présentation du Cirad 3

1.1.2 Le Cirad à La Réunion..... 3

1.1.3 L'île de La Réunion 4

1.2 Place du stage dans le projet..... 5

2. Création de la BDD..... 7

2.1 Données disponibles 7

2.2 Constat 9

2.3 Les choix technologiques 10

2.4 Conception de la BDD 11

2.4.1 Le Modèle Conceptuel de Données..... 11

2.4.2 Le Modèle Logique de Données 13

2.4.3 Le Modèle Physique de Données..... 14

2.5 Création des tables 15

2.6 Remplissage, explorations et exportations des données..... 15

CONCLUSION.....16

BIBLIOGRAPHIE

TABLE DES SIGLES

ANNEXES

REMERCIEMENTS

Je tiens d'abord à remercier mon maître de stage, Christophe Poser, pour son encadrement, son soutien et sa disponibilité sans limite. Egalement pour ses précieux conseils et sa confiance, mais aussi pour sa bonne humeur permanente et sa passion si communicative.

Par la suite, je voudrais remercier Frédéric Chiroleu et Jérôme Queste, de m'avoir aidé fréquemment dans l'accomplissement de mon stage.

Enfin, je tiens à remercier toute l'équipe du Cirad de la station de La Bretagne : les chercheurs Jean-François Martiné, José Martin et Damien Sabatier du pôle Canne à sucre, les techniciens Maurice Gueno et Patrice Leblé, le personnel administratif, les stagiaires, ainsi que Thomas Dumond de l'équipe d'eRcane. Ils ont réussi à instaurer une atmosphère chaleureuse et vivante, et ont contribué à faire de mon séjour à La Réunion une expérience enrichissante et inoubliable.

RESUME

La culture de la canne à sucre est un véritable enjeu économique, social et environnemental sur l'île de La Réunion. La surface plantée en canne à sucre s'estime à 26% dans les « Hauts », où les températures sont les plus basses, engendrant de nombreuses contraintes culturelles.

C'est pourquoi le Cirad réalise des essais sur des variétés nouvellement adaptées à ces contraintes de basses températures. Ces essais se répartissent sur quatre stations : Bassin Plat, La Mare, Ligne Paradis et les Colimaçons et génèrent un flot important de données agronomiques, actuellement enregistrées dans des feuilles de calculs Excel. Les données collectées contribuent par la suite à des études statistiques.

Les chercheurs du Cirad des pôles Sud et Nord de l'île souhaitent mettre en commun leurs données dans une base, exploitable via le Web. Une structure de cette base a été proposée après un travail de réflexion préalable reposant sur la conception de modèles.

Mots clés : canne à sucre, La Réunion, les « Hauts », basses températures, contraintes culturelles, base, structure, conception, modèles.

ABSTRACT

On the Reunion Island, Sugarcane growing represents real economic, social and environmental issues. The area planted with sugar cane is evaluated to 26% on the highest places called "The Top" where the temperature is very low and causing many farming constraints.

That's why Cirad makes tests on sugarcane varieties which are relatively resistant to these low temperature constraints. These tests are distributed on 4 stations: "Bassin Plat", "La Mare", "Ligne Paradis" and "Les Colimaçons" and generate an important agronomical data currently saved on Excel SpreadSheets. The data contribute then to statistic studies.

Cirad Researchers from the South and the North of the Island would like to share their data, put it in a database on the web. After a models conception working, a database structure has been suggested...

Keywords: sugarcane, Reunion, "The Top", low temperatures, farming constraints, database, structure, models, conception.

INTRODUCTION

Depuis plusieurs siècles, la canne à sucre est un des moteurs de l'économie réunionnaise. Agro-industrie, transport, énergie, recherche, tourisme : elle est présente dans la plupart des secteurs d'activité de l'île et génère ainsi 12 000 emplois. En 2009, la filière canne à sucre représente plusieurs millions d'euros de chiffre d'affaires, avec une production d'environ 1 900 000 tonnes de cannes.

A la Réunion, la canne occupe plus de la moitié des surfaces cultivées avec près de 25 000ha (1ha = 10 000m²). La culture de la canne à sucre occupe une place importante dans l'économie de l'île puisque les produits de sa transformation représentent 85% des exportations de l'île. En effet, chaque année la quantité de sucre issue de la canne à sucre s'élève à 200 000t. Mais l'utilisation de la canne à sucre ne se limite pas à la production de sucre : la canne à sucre est également une source d'énergie renouvelable. A sa livraison à l'usine la canne est séparée en deux composants : le jus duquel sera extrait le sucre, et la bagasse utilisée comme combustible. Les deux centrales thermiques mixte charbon/bagasse de l'île permettent de produire 277 GWH* d'électricité. La bagasse produit 10% de l'électricité consommée sur l'île, elle représente un tiers des énergies renouvelables de La Réunion.

Un autre constituant de la production de canne à sucre, la mélasse, produit final non cristallisé, visqueux est un mélange d'eau, de sucre et de composés minéraux hautement fermentescible. Elle sert pour la production de levures de boulangerie et de brasserie, pour l'alimentation animale ou peut encore être distillée pour produire de l'alcool d'origine agricole : le rhum industriel, « traditionnel » (à l'opposé du rhum agricole réalisé à partir du jus).

Enfin, la canne à sucre a un impact environnemental positif : elle joue un rôle dans la gestion du territoire réunionnais en occupant 10% du territoire et façonnant des paysages liés à l'image attractive de l'île pour le tourisme.

Cependant, les surfaces agricoles se réduisent par les constructions urbaines liées à la croissance démographique de l'île. Avec une population de plus de 800 000 habitants en 2008, l'INSEE estime que la population pourrait atteindre plus d'un million d'habitants d'ici 2030. Par ailleurs, le prix d'exportation du sucre est à la baisse depuis la réforme de l'OCM* et confronte les filières canne dans les DOM* à recourir, pour le moment à des compensations de l'Etat. Une des solutions pour palier le manque de terres disponibles et ne pas risquer de sous utiliser les usines existantes consiste à produire plus de canne à sucre sur la zone des « Hauts » de l'île. La diversité des sols et les micros climats représentent cependant, sur cette zone, des réelles contraintes. Deux organismes de recherche collaborent sur le sujet. Les agronomes sélectionneurs d'eRcane créent de nouvelles variétés de canne à sucre plus appropriées aux différents environnements pédoclimatiques de l'île.

Le Cirad* intervient dans ce cadre, en accompagnant la sélection par des essais en phytopathologie et en agronomie.

De nouvelles variétés mieux adaptées aux conditions de cultures sous contraintes ont été sélectionnées et diffusées aux planteurs. Des essais se poursuivent pour comprendre les phénomènes d'adaptation, le stage est articulé sur l'archivage des données pour une exploitation ultérieure simplifiée.

1. Contexte

1.1 Le Cirad et l'île de La Réunion

1.1.1 Présentation du Cirad

Le Cirad est l'institut français de recherche agronomique pour le développement des pays du Sud et des Doms-Toms. Créé en 1984, cet organisme est sous la tutelle des ministères français chargés de la Recherche et des Affaires étrangères. C'est un Etablissement Public à Caractère Industriel et Commercial (EPIC*).

Les actions du Cirad s'inscrivent dans une démarche de développement durable collaborative, c'est-à-dire qu'elles prennent en compte les impacts économiques, écologiques et sociaux des sociétés et des territoires des pays du Sud. Le Cirad produit et transmet de nouvelles connaissances pour accompagner le développement agricole et contribuer au débat sur les grands enjeux mondiaux de l'agronomie. En France comme à l'étranger, il mène des recherches et des expérimentations, ainsi que des actions de formation et d'innovation, principalement dans le secteur agricole, forestier et agro-alimentaire.

Le Cirad possède 25 dispositifs de recherche en partenariat dans le monde et 7 Pôles scientifiques à vocation régionale dans l'outre-mer français (Cirad 2010).

1.1.2 Le Cirad à La Réunion

Le Cirad de La Réunion est le plus important dispositif de recherche hors de l'Hexagone. Il a pour objectif principal de proposer des solutions innovantes aux problèmes que rencontre l'agriculture réunionnaise. Suite à la mise en place des programmes opérationnels (PO) de l'Union Européenne (UE) pour la période 2008-2013, le Cirad à La Réunion se voit attribuer une triple mission :

- Produire des résultats scientifiques d'excellence
- Répondre aux besoins du développement rural de l'île
- Mener une politique de coopération régionale active dans l'océan Indien

Les laboratoires, les serres et les bureaux du Cirad à la Réunion sont localisés sur les deux principaux sites de l'île : à La Bretagne et à Ligne Paradis tandis que les 35 Ha de terrains d'expérimentations se répartissent sur quatre stations : Le Bassin Plat, La Mare, Ligne Paradis et Les Colimaçons. Au total, 181 salariés permanents, dont 55 chercheurs et 120 techniciens VCAT* et autant de doctorants et de stagiaires.

Le Cirad à la Réunion intervient principalement dans six domaines d'activités :

- la culture de la canne à sucre
- les cultures maraichères et les plantes aromatiques
- l'agriculture durable
- l'environnement et la forêt
- l'élevage
- la protection des plantes et l'agro-alimentaire

Depuis le 1^{er} Janvier 2007, l'institut est structuré en 3 pôles de recherche :

- Qualité des productions agricoles et alimentaires tropicales (pôle KAPPA)
- Risque environnementaux, agriculture et gestion intégrée des ressources (pôle REAGIR)
- Protection des plantes (3P)

Chaque pôle est articulé autour de plusieurs programmes de recherche (12 sont planifiés pour la saison 2008-2013).

Un des programmes du pôle REAGIR concerne la conception de systèmes de production performants et durables pour les cultures annuelles, tant du point de vue économique que social et environnemental. Il est assuré par l'UPR* n° 102 Systèmes de culture annuels qui a affecté 6 chercheurs sur place.

« A la lumière des crises récentes qui ont touché les cours des produits agricoles, l'usage des ressources, la gestion de l'environnement, il devient essentiel d'imaginer des manières de produire, performantes certes, mais aussi plus propres, plus économes, diversifiées et résilientes, pour permettre aux paysans du Sud de vivre mieux dans le monde de demain. » (ref. site internet Cirad)

C'est sur cette problématique que les 6 chercheurs aux domaines de compétences diversifiés œuvrent.

1.1.3 L'île de La Réunion

L'île de la Réunion se situe dans l'océan Indien, à 200 km au sud-ouest de l'île Maurice et 700 km à l'Est de Madagascar. Avec une superficie de 2 512km², ce département français d'outre-mer est la plus grande île de l'archipel des Mascareignes.

Cette île volcanique s'est formée progressivement depuis 2 millions d'années. Elle abrite un volcan encore actif : le piton de la Fournaise (2 631m) et son sommet est le piton des Neiges qui culmine à 3 071m. Sa géographie est extrêmement contrastée d'un point de vue climatique, orographique et pédologique (*cf. annexe 1*).

En 2009, la démographie du département était regroupée dans les villes de la plaine côtière, et tout particulièrement dans la capitale régionale et le chef-lieu du département : Saint Denis. Fortement métissée de part son histoire (colonisation, esclavage, immigrations spontanées), la population de la Réunion a des origines des quatre coins du monde : européenne, malgaches, africaines (« Cafres »), indiennes et chinoises.

L'économie repose sur le tourisme et la culture de cannes à sucre. Avec près de 400 000 visiteurs chaque année, le secteur tertiaire est devenu le principal employeur de l'île. Mais le taux de chômage du département reste élevé avec 24,5% en 2008, en partie à cause du manque d'industrialisation.

L'agriculture réunionnaise réunit encore près de 4% de la population active pour une surface agricole utile SAU* de 47 389 Ha. La canne représente plus de 60% de la SAU (seulement 26 % sont consacrés aux Hauts de l'île). Mais cette SAU diminue à cause du déclassement des terres agricoles pour les constructions urbaines.

1.2 Place du stage dans le projet

Comme nous l'avons vu, les agriculteurs tentent de cultiver la canne à sucre dans « Les Hauts » de l'île. Seulement, la zone communément appelée « Hauts » de La Réunion (à partir d'une altitude de 200 m pour la côte Est et de 600 m pour la côte Ouest, cf. annexe 2) révèle des contraintes climatiques et organisationnelles. Les températures sont notamment les plus basses que dans le reste de l'île, ce qui ralentit considérablement la pousse de la canne et rend cette dernière vulnérable par rapport aux mauvaises herbes.

La recherche intervient localement pour ces zones aux contraintes notamment en proposant aux planteurs des variétés adaptées aux contraintes climatiques de chaque zone afin d'obtenir une meilleure qualité technologique¹ de la canne. Le CIRAD accompagne cette démarche et travaille en collaboration avec le centre de sélection variétale réunionnais eRcane.

Autrefois connu sous le nom de Centre d'Essai, de Recherche et de Formation (CERF*), eRcane qui choisit pour signature « Valoriser la ressource canne », se positionne comme un centre de recherche tourné vers l'avenir, en quête de productivité sucrière et des nouvelles valorisations de la canne à sucre. Parmi ses activités, eRcane a pour mission la création variétale. Cette création variétale se

¹ La qualité technologique d'une canne est habituellement estimée par les industriels en fonction de sa richesse en sucre, de son aptitude au broyage, de la pureté de son jus, de la bonne cristallisation de son sucre.

déroule en deux étapes (hybridation et sélection) et s'appuie sur des recherches génétiques.

Le principe de l'hybridation consiste à croiser deux à deux les fleurs de variétés de canne possédant des qualités recherchées (richesse en sucre, fort rendement, aptitude à la repousse, résistance aux maladies, production importante de biomasse...).

Pour sélectionner ses nouvelles variétés, eRcane dispose d'une importante collection variétale provenant du monde entier, de pépinières, de serres d'hybridation et de six stations d'essais réparties dans toute l'île. La diversité des sols et des micro-climats réunionnais oblige à créer des cannes adaptées à des conditions très différentes, à La Réunion mais aussi à d'autres zones de production. Pour exemple, on connaît les variétés suivantes :

- R 570 : toutes zones, sauf altitudes supérieures à 550 m dans l'Ouest et le Sud, ou supérieures à 300 m dans l'Est. (c'est la canne la plus plantée, elle est d'ailleurs exportée dans plusieurs pays)
- R 575 : toutes les zones (canne de début de campagne).
- R 577 : l'Ouest et le Sud de l'île, de 400 à 700 m (canne de milieu de campagne).
- R 583 : zones d'altitude sèches et non irriguées de la côte Sous-le-vent.
- R 585 : zones difficiles de l'Est et du Sud.

Chaque parcelle expérimentale est intégralement pesée pour déterminer le tonnage/hectare, et des échantillons de tige de cannes sont analysés au laboratoire d'eRcane pour déterminer le taux de fibre, la richesse en sucre, la pureté. Ces paramètres sont d'ailleurs utilisés pour déterminer le prix de chaque livraison réalisée par les planteurs qui livrent à l'usine leur production.

Parallèlement, des essais maladies sont mis en place en collaboration avec le service pathologie du Cirad afin d'évaluer la résistance de ces variétés aux principales maladies présentes à La Réunion (Charbon, Echaudure des feuilles, Gommose) ainsi qu'appréhender leur résistance aux insectes.

Au final, il s'agit de reconnaître, dans chaque zone, les cannes susceptibles d'égaler ou de dépasser les meilleurs standards localement utilisés par les agriculteurs. Le Cirad intervient également dans le domaine de l'agronomie en étudiant les moyens d'améliorer les itinéraires techniques, (densité de plantation, gestion de l'entreligne, amélioration de la fertilisation...) ainsi que les capacités d'adaptation d'une variété à un environnement contraint : « Les Hauts ».

Il est important de comprendre que tous ces tests de sélection génèrent des données qui doivent être enregistrées, structurées dans une base de données, exploitable par toute personne requérant des données, qu'elle que soit sa fonction.

Le but de mon stage est de proposer une structure de base dans laquelle s'organisent les données de mon maître de stage, Christophe Poser, qui travaille justement sur la contrainte de la culture de la canne dans les «Hauts». Ce dernier étudie plus précisément l'aptitude pour une variété à s'adapter à la contrainte de basse température relative présente sur les zones d'altitude. Il a réalisé avec son équipe des essais à partir de boutures de différentes variétés contrastées de canne à sucre en suivant les différents processus physiologiques (levée, tallage, émission de feuilles, élongation des tiges ...) depuis la germination jusqu'à la récolte. Son objectif est de montrer que certaines variétés adaptées aux conditions des Hauts le sont grâce à de meilleures performances comparées aux variétés classiques. Les données biologiques recueillies doivent être traitées séparément selon le processus étudié et mis en relation avec des données climatiques correspondantes principalement articulées autour des données de température relevées sur chacun des sites d'expérimentation.

Il est souhaitable que cette structure de base de données soit compatible pour ses collègues qui possèdent le même type de données et que des formats d'échanges, de saisies, de consultations soient compatibles avec un travail sur un serveur centralisé accessible via le Net. Les données sont actuellement disponibles sur des feuilles Excel, la plupart servant à exploiter statistiquement ces données, avec le logiciel R.

Cette base de données doit avoir avant tout un côté pratique pour obtenir facilement les données que l'on cherche, sans replonger dans les feuilles de calculs d'Excel.

2. Création de la BDD*

2.1 Données disponibles

Les données présentes se répartissent en huit essais réalisés par l'équipe. Les essais se sont déroulés sur une période de 3 ans, sur quatre sites différents, depuis juillet 2008, à partir de boutures d'un seul bourgeon ou « œilleton » issu de tige de cannes prélevés dans des champs identifiés ayant leur propre histoire (date de plantation, climat...). On peut observer en *annexe 3* des photos de boutures. Des mesures biométriques sont réalisées sur ces boutures (poids, taille, richesse en sucre, longueur, densité sèche...) dont le temps de germination sera lui aussi mesuré dans des conditions thermiques et hygrométriques suivies en conditions naturelles (au champ) ou en enceinte climatiques régulées.

Outre le processus de germination ou levée, d'autres processus sont suivis. Les mesures relatives à la vitesse d'apparition de tige, de feuilles sont effectuées. Les bourgeons de chaque bouture ont donc levé, poussé, produit plusieurs tiges sur lesquelles apparaissent des feuilles. A pas de temps plus ou moins régulier, les nombres

de feuilles et de tiges en présence sur ces touffes, issues des bourgeons primaires sont relevés.

Des mesures de biomasse peuvent également avoir lieu selon les essais. Ils consistent à répertorier les composantes de la masse végétale en partition de feuilles vertes ou sèches, parties de tige « usinables », (contenant le jus sucré) ou les sommets de tiges, ou « choux » non exploités en sucrerie.

L'évolution de la croissance du feuillage, révélateur de la conquête de l'espace entre lignes de cannes à sucre par la variété est elle aussi étudiée. Des mesures de radiations photosynthétiques active (PAR) entre 400 et 700 nm sont effectuées à l'aide d'un analyseur de couvert végétal Accupar LP 80.



Photo du ceptomètre Accupar LP 80

Une règle de 86.5 cm de long mesure par l'intermédiaire de 8 capteurs une différence entre le rayonnement direct au dessus et au dessous du couvert végétal. Une mesure d'interception en pourcentage est ainsi obtenue depuis la levée jusqu'à la fermeture du couvert définie selon la bibliographie autour de 70 %.

A la récolte encore, une biomasse peut être effectuée à l'échelle de la parcelle, du bloc, de la placette ou bien encore à l'échelle de la touffe produite à partir de la bouture d'origine connue ayant produit des tiges ; le nombre de tiges produites par bouture ainsi que le poids total produit pourra lui aussi être mesuré.

Le principe d'essais agronomiques repose sur le principe de mesures effectuées en répétition sur des blocs supposés identiques représentant des répétitions.

2.2 Constat

Il s'avère que l'ensemble des essais est effectué selon un cheminement réfléchi, correspondant à des besoins de démonstrations scientifiques rigoureuses répondant à la mise en évidence d'une partie des paramètres à expliquer. Tous ces paramètres ne pouvant être étudiés simultanément, l'ensemble des mesures n'est pas effectué sur la totalité des essais.

La difficulté dans l'établissement de la base réside à ce niveau dans le fait que les champs renseignés ne seront pas toujours complets pour tous les essais bien que la base doive impérativement intégrer l'ensemble des paramètres mesurés, plus ceux éventuellement calculés à partir des précédents. La base devra, de plus, rester d'une structure logique et cohérente.

Une autre difficulté réside dans le fait d'associer à ces mesures ponctuelles destructives (biomasse, qui nécessite de couper les cannes) ou répétitives (longueur de tiges) des pas de temps plus ou moins réguliers. La notion de temps pour des phénomènes d'ordre physiologique est exprimée communément pour les plantes par une « somme de degré jour » correspondant à l'accumulation de températures moyennes journalières², le plus souvent référencé à une température de base issues de données météorologiques provenant de fichiers annexes.

Les études visent principalement à comparer le comportement de variétés les unes par rapport aux autres. Il convient alors de renseigner le mieux possible l'origine de la variété (parents desquels est issu le croisement) ainsi que les qualités éventuelles connues de la variété (résistance aux maladies, aux insectes, type de port droit ou versant, attitude végétative, couleur de certains organes, etc.) qui constituent les caractéristiques propres de la variété et permettent de la reconnaître au champ.

Enfin, certains paramètres comme la provenance des boutures souches, qualité du sol utilisé, type de conduite (irrigué ou non), devons eux aussi être notifiés dans la base.

L'ensemble de ces contraintes dans l'établissement de la base a été pris en compte et a nécessité de nombreux questionnements aux différents acteurs des expérimentations et la participation aux travaux fastidieux de mesures tant durant la culture qu'au moment de la récolte. L'intégration du travail de saisie au champ fut également une pièce maitresse dans la compréhension du fonctionnement de la plante et dans la compréhension des phénomènes mis en jeu.

L'analyse suivante découle de la volonté de prendre en compte de façon exhaustive l'ensemble des données présentes et susceptibles de générer des fichiers permettant l'analyse statistique des données en présence.

² Moyenne horaire ou moyenne agricole : $(\text{Min} + \text{Max}) / 2$

2.3 Les choix technologiques

Avant de commencer la construction de la base de données, il est nécessaire de faire les bons choix technologiques, à savoir quel est le système de gestion de base de données (SGBD*), le système d'exploitation utilisé, quel serveur web et quel serveur d'application va-t-on utiliser.

Le SGBD est le cœur des bases de données. Il existe trois types de bases de données : les bases de données relationnelles, les bases de données purement objet et les bases de données relationnel-objet. Dans le cadre d'étude, il s'agit d'un SGBDR*.

Le système d'exploitation le plus couramment utilisé au Cirad est Windows, mais le système d'exploitation importe peu pour le choix des serveurs web et d'applications.

Un serveur web est un logiciel permettant à des clients d'accéder à des pages web, c'est-à-dire en réalité des fichiers au format HTML* à partir d'un navigateur (aussi appelé browser) installé sur leur ordinateur distant. Un serveur web est donc un « simple » logiciel capable d'interpréter les requêtes HTTP* arrivant sur le port associé au protocole HTTP (par défaut le port 80), et de fournir une réponse avec ce même protocole.

A ce jour, le serveur web sera le serveur Apache. Outre sa gratuité et sa disponibilité, Apache a depuis longtemps fait les preuves de sa robustesse, de sa rapidité et de sa stabilité. Qui plus est, il offre une sécurité maximale sous Windows comme sous Unix.

Enfin, le serveur applicatif sera PHP*. Nous utiliserons donc un serveur web Apache, le serveur applicatif Perl, ainsi qu'une base de données MySQL*, qui permet de créer des tables contenant plusieurs milliards d'enregistrements. Ces logiciels sont gratuits, très performants et utilisés dans de très nombreux projets de base de données.

Pour le maintien et l'exploitation de la base, un autre outil va être indispensable. Cet outil est *PhpMyAdmin*, c'est une application Web de gestion pour les systèmes de gestion de base de données MySQL réalisée en PHP et distribuée sous licence GNU GPL.

Cette interface graphique permet d'exécuter, très facilement et sans grandes connaissances dans le domaine des bases de données, de nombreuses requêtes comme les créations de table de données, les insertions, les mises à jour, les suppressions, les modifications de structure de la base de données. Ce système est très pratique pour sauvegarder une base de données sous forme de fichier de type « .sql » et ainsi transférer facilement ses données. De plus celui-ci accepte la formulation de requêtes directement en langage SQL, cela permet de tester ses requêtes.

2.4 Conception de la BDD

La conception d'une BDD nécessite un travail de réflexion préalable qui repose sur trois grandes étapes :

- Établir un MCD*
- Établir un MLD*
- Établir un MPD*

2.4.1 Le Modèle Conceptuel de Données

Le MCD, ou modèle entité-association fournit une description graphique pour représenter les données sous forme de diagrammes contenant des entités et des associations. Les entités ou objets sont par exemple des variétés de canne à sucre, des mesures, des essais, etc. Des entités de même type (variétés R 570, R 577, R 583, etc.) forment un ensemble appelé classe d'entité. Les classes d'entités possèdent des propriétés appelées « attributs ». Un attribut de l'entité variété de canne à sucre sera le nom de la variété. Les classes d'entités sont représentées par des rectangles partagés en deux parties : la partie du haut comprend le nom de l'entité, la partie du bas les attributs qui caractérisent l'entité. Chaque individu d'une entité doit être identifiable de manière unique. C'est pourquoi toutes les entités doivent posséder un attribut ne prenant pas deux fois la même valeur, on dit qu'il est sans doublon. Il s'agit de l'ID* qui par convention est souligné.

Une association, plus communément appelée « relation » représente les liens sémantiques qui peuvent exister entre plusieurs entités tels que « contient » entre l'entité « site » et l'entité « parcelle ». Ces associations sont représentées par des ellipses ou des losanges. Une association peut être dépourvue d'attribut.

Après mûre réflexion sur les contraintes qu'engendre l'ensemble des essais à répertorier dans la base de données (cf. 2.2), voici le MCD proposé (cf. *Annexe 4*).

La relation d'agrégation entre les tables « SITE », « PARCELLE », « BLOC » et « PLACETTE » informe sur l'aspect géographique de l'essai. C'est précisément sur une placette d'un bloc d'une parcelle d'un site que l'on fait des observations. Les données météorologiques (Précipitation, Température et Rayonnement) sont relevées au niveau de la parcelle à l'aide d'une centrale d'acquisition de type Campbell Scientific sur laquelle sont reliés des capteurs.

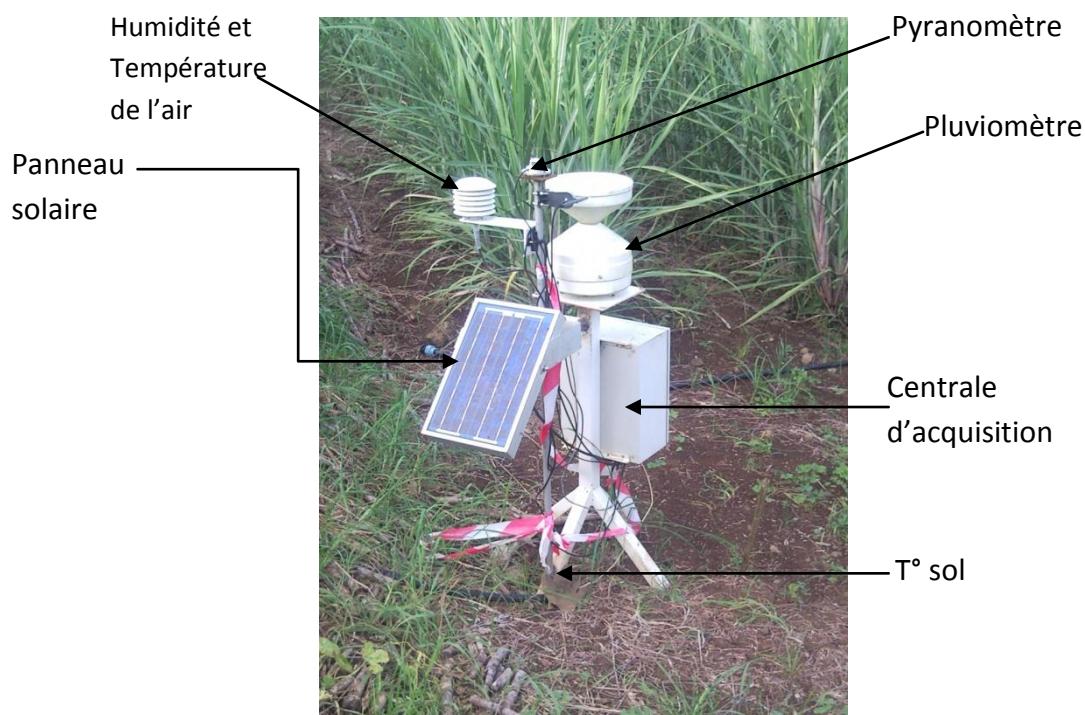


Photo d'une centrale d'acquisition de type Campbell Scientific

D'après Jérôme QUESTE, chercheur au Cirad, anciennement informaticien, cette BDD est atypique. Elle tire son originalité grâce à sa table « PLANT ». En effet, cette table regroupe les ID des tables qui sont reliées à elle pour les concaténer de manière à n'en faire qu'un ID unique qui correspond à « l'objet » mesuré, observé. Par conséquent, il est possible de regrouper tous les essais dans une même structure de BDD, ce qui n'était pas si évident à première vue. En effet, les champs de saisie ne sont pas tous présents d'un essai à l'autre. Par exemple, pour l'essai *Germiléo*, les mesures sont faites sur des variétés plantées sur des blocs alors que pour l'essai *Position 1*, les mesures sont faites sur des bourgeons de variétés, plantées dans des chambres climatiques, on a donc pas de notion de blocs et de placettes pour l'essai *Position 1*.

Cette différence de cheminement pour arriver à l'observation que l'on veut faire s'avère être une réelle contrainte dans l'organisation d'une BDD, c'est la raison pour laquelle dans le MCD, la table « PLANT » est reliée aussi bien à la table « PARCELLE » ou la table « BLOC » que la table « PLACETTE ». Ainsi, l'ID concaténé se présenterait sous la forme *SITE-PARCELLE-BLOC-PLACETTE-ESSAI-VARIETE-TIGE-POSITION* où figurerait une valeur arbitraire nulle, un « # » par exemple, lorsqu'il n'y aurait pas la notion d'un des éléments.

Il faut noter que toutes les tables comprendront un ID qui ne sera pas seulement numérique mais contiendra les noms abrégés des enregistrements : pour le site « La Bretagne », l'ID ne sera pas désigné par un « 1 » mais par « BRE ». Ce qui rend la lecture de l'ID unique de la table « PLANT » beaucoup plus aisée. On aurait donc par exemple, pour l'essai *Germiléo* un enregistrement avec un ID tel que « COL-CHP1-B2-#-Germ-R570-

»³ et pour l'essai *Position 1* un enregistrement avec un ID tel que « BRE-CBR18-##-Pos1-R570-T2-P5 »⁴.

Les tables « ESSAI », « VARIETE », « TIGE » et « POSITION » vont bien sûr s'associer à la table « PLANT » pour savoir, suivant les essais recherchés, à quel bourgeon (qui se situe à une position précise sur la canne) de quelle tige de quelle variété se réfère la mesure observée. La table « PLANT » ne contient donc rien d'autre que les ID de toutes les tables annexes, sauf celui de la table « OBSERVATION » et celui de la table « TEMP ». La collecte des températures se fera parallèlement à l'aide du site et de la parcelle, à une date précise.

Enfin, l'ID généré par la concaténation des autres ID de la table « PLANT » va être transmis à la table « OBSERVATION ». Les attributs de cette table ne sont autres que toutes les mesures établies dans tous les essais, précédées d'un attribut date qui permettra d'associer un pas de temps aux mesures ponctuelles destructives ou répétitives comme souhaité.

A présent que le MCD est établi, il faut passer à la deuxième étape : l'établissement du MLD.

2.4.2 Le Modèle Logique de Données (cf. Annexe 5).

L'étape MLD se situe chronologiquement juste après l'étape MCD et revient à présenter les objets du MCD en remplaçant les associations entre entités par des cardinalités. La cardinalité d'un lien entre une entité et une association précise le minimum (0 ou 1) et le maximum (1 ou n) de fois qu'un individu de l'entité peut être concerné par l'association. Un site contient obligatoirement une à plusieurs parcelles, on a donc une cardinalité de 1 :n. Par contre, une parcelle *p* n'appartient qu'à un seul et unique site, on a donc une cardinalité de 1 :1. C'est à peu près le même principe pour la relation entre la parcelle et le bloc : Un bloc appartient forcément à une seule et unique parcelle, on a donc une relation de cardinalité de 1 :1. Cependant, comme on l'a exposé précédemment, une parcelle n'est pas toujours partagée en blocs, on a donc une cardinalité de 0 :n. On en conclue que la cardinalité minimale de 1 se justifie par le fait que l'individu de l'entité liée à besoin de l'association pour exister : un bloc n'existe pas sans une parcelle.

³ Col = Colimaçons, CHP1 = Champ1, B2 = Bloc2, Germ= Germiléo, R570 = Variété de canne R 570

⁴ Bre = Bretagne, CBR18 = Chambre Climatique à 18°, Pos1=Position 1, T2 = Tige2, P5 = Position5 (Bourgeon de rang n° 5)

2.4.3 Le Modèle Physique de Données

Le Modèle Physique de Données revient à présenter les objets du MLD sous une forme compréhensible par un SGBD : les objets représentés sont désormais des tables et les liens qui les unissent. La traduction du MLD conduit à un MPD (*cf. Annexe 6*) contenant tout les attributs et qui précise son type, sa nature :

- Varchar : Chaîne de caractères
- Int ou Integer : Chiffre entier positif
- Timestamp : Date au format JJ/MM/AA et heure format HH:MM
- Float : Chiffre flottant, à virgule

On peut également voir les éléments « PK » et « #FK », ce qui signifie respectivement « PRIMARY KEY », clé primaire et « FOREIGN KEY », clé étrangère. C'est par convention que l'on souligne les clés primaires et que l'on place un « # » devant une clé étrangère.

Les enregistrements d'une table doivent être uniques, cela signifie qu'au moins une colonne doit servir à les identifier. Les attributs possédants ces indices sont souvent les ID des tables. L'élément « PK » indique qu'il ne peut pas y avoir d'enregistrements ayant la même valeur pour l'attribut dont il est question : Dans la table « SITE », l'ID_SITE est une clé primaire. On ne peut donc pas avoir deux fois l'ID_SITE « BRE » pour le site « La Bretagne ». Une clé étrangère fait référence au contenu d'un attribut d'une autre table : Dans la table « TEMP », l'ID_PARC est une clé étrangère qui fait référence à l'ID_PARC de la table « PARCELLE ». Ces deux ID_PARC contiennent les mêmes informations mais sont dans deux tables différentes. A savoir qu'un attribut peut à la fois être une clé primaire et une clé étrangère (enregistrement unique qui est utilisé dans une autre table), c'est le cas de cet ID_PARC.

N.B : Une table peut avoir plusieurs clés étrangères mais une seule clé primaire (qui peut être composée de plusieurs colonnes : c'est pour cela qu'il y a plusieurs attributs soulignés dans les tables « TEMP » et « OBSERVATIONS »).

Le problème avec les clés primaires et clés étrangères est qu'au moment de la mise à jour de la BDD (remplissage, suppression, etc.) le SGBDR va vérifier que chaque clé étrangère prenne bien une des valeurs de la ou des colonne(s) qu'elle référence. Ce phénomène garantit l'intégrité référentielle des données. Les clés engendrent une certaine rigidité et une rigueur que l'on ne peut pas avoir dans ce cadre d'étude. Ce qui explique l'absence de clé étrangères dans certaines tables.

En général quand on regarde le MPD, on est aussitôt capable de construire la BDD en langage SQL. Pour la création des tables, on ne codera pas directement sous la console, on utilisera *PhpMyAdmin*.

2.5 Création des tables

Avant de créer les tables, il faut bien entendu créer la base de données qui va recueillir toutes les tables du MPD. La base de données, nommée dbCirad, va apparaître dans la liste des bases de données (*cf. Annexe 7*).

Une fois la base créée, on nous indique qu'aucune table n'existe, il faut donc créer une à une nos tables dans le champ "Créer une nouvelle table sur la base dbCirad" en indiquant le nom de la table et son nombre de champs (le nombre d'attributs) (*cf. Annexe 8*). La suite consiste à nommer chaque attribut : pour la table « SITE » on aura donc 3 champs nommés « ID_SITE », « NOM_SITE », « INFO », et indiquer le type des attributs. Il faudra répéter l'opération pour les 11 tables à créer.

2.6 Remplissage, explorations et exportations des données

Une fois toutes les tables renseignées, *PhpMyAdmin* propose un onglet « Importation » qui permet d'importer un fichier à différents formats, notamment les fichiers « .sql ». Le fichier pourra contenir les requêtes de remplissage des tables qui s'exécutera lors de l'importation du fichier.

PhpMyAdmin possède également un onglet « SQL » pour l'exécution tout type de requêtes SQL permettant d'explorer les données de la base.

Enfin, il est possible d'exporter la base (grâce à l'onglet « Exporter »). On va pouvoir récupérer la base de données sous forme d'un fichier texte « .sql » contenant toutes les requêtes SQL qu'il aura fallu pour créer et remplir la base. Ce fichier servira à :

- transmettre la base de données sur Internet : Le fichier « .sql » généré permettra de reconstruire la base de données grâce à l'outil d'importation de *PhpMyAdmin* sur le serveur Web choisi.
- la sauvegarde de la base de données : Il est important de conserver une sauvegarde de la base de données.

Avec ce même onglet « Exporter », il sera possible d'exporter au format .CSV les résultats de requêtes SQL. Ce fichier pourra être directement travaillé sous un tableur Excel ou le logiciel R.

CONCLUSION

Grâce à ce stage de fin d'études, j'ai pu comprendre ce qu'était la Bioinformatique. En effet durant ces 5 mois de stage, j'ai pu me rendre sur les terrains de champs de canne à sucre avec les techniciens, voir la composition de la canne, prendre connaissance des essais réalisés. Je me suis imprégné du vocabulaire utilisé en Agronomie en discutant avec différents interlocuteurs dans le but de répondre à la problématique d'une base de données. J'ai vraiment joué le rôle du Bioinformaticien, celui qui fait le lien entre deux mondes, si proches mais pourtant différents : l'Agronomie et l'Informatique. Analyser les données n'a pas toujours été très évident mais la résolution du problème posé m'a permis d'acquérir une bonne expérience, et de réaliser que la mise en pratique des connaissances dans le monde du travail n'est pas toujours aussi évidente que les exercices théoriques proposés dans le cadre de la formation.

Ayant eu un grand intérêt pour l'essai *Germiléo*, outre la prise de notes, j'ai décidé de porter un regard plus profond sur ces données à travers une étude statistique complète d'un processus. Cette étude s'est réalisée grâce au logiciel R. Les connaissances que j'avais à propos de ce logiciel, utilisé durant la formation, ne suffisaient pas pour aboutir. Par l'intermédiaire d'une formation et du statisticien Frédéric Chiroleu, que je remercie encore, j'ai pu obtenir un niveau satisfaisant en programmation R et j'ai ainsi pu terminer le script de mon étude.

Au final, ce stage reste une expérience enrichissante et inoubliable. Ce que j'ai pu accomplir au cours de ces 5 mois me donnent l'envie de continuer mes études dans le domaine de la Bioinformatique.

BIBLIOGRAPHIE

CEDUS, Centre d'Etudes et de Documentation Du Sucre, *Sucre et autres débouchés : Mémo statistique*, 75116 Paris, Mai 2011. 33 p.

Courteau, Anaïs. *La canne à sucre et l'environnement à La Réunion*. 48 p. Rapport de stage de Maîtrise : IUT Professionnalisé Génie des Territoires et de l'Environnement Filière Diagnostic et Gestion des Systèmes Ecologiques : Université de Franche-Comté ; Cirad : 2005

Huet, Camille. *Analyse expérimentale de la morphogénèse de deux génotypes de canne à sucre jusqu'à leur maturité « agronomique »*. 28 p. Rapport de stage de Maîtrise : AgroParisTech : Institut des Sciences et Industries du vivant et de l'Environnement ; Cirad : 2009

Roguet, Stéphanie. *Etude du facteur température dans l'adaptation de variétés de canne à sucre dans « les Hauts » de La Réunion*. 77 p. Mémoire fin d'études : ISTOM : Ecole d'Ingénieur en Agro-Développement International : Cergy-Pontoise ; Cirad : 2010

Cirad : La recherche agronomique pour le développement. *Le Cirad, en bref*, [En ligne]. <<http://www.cirad.fr/qui-sommes-nous/le-cirad-en-bref>> (consultation le 20-07-2011)

Cirad : La recherche agronomique pour le développement. *La Réunion et Mayotte*, [En ligne]. <http://www.cirad.fr/reunion/cirad_reunion> (consultation le 20-07-2011)

Cirad : La recherche agronomique pour le développement. *Unités de Recherche*, [En ligne]. <http://reunion-mayotte.cirad.fr/recherche_en_partenariat/unites_de_recherche> (consultation le 20-07-2011)

Cirad : La recherche agronomique pour le développement. *Systèmes de culture annuels*, [En ligne]. <<http://www.cirad.fr/nos-recherches/unites-de-recherche/systemes-de-culture-annuels>> (consultation le 20-07-2011)

Cirad : Unité de recherche. *Systèmes canniens*, [En ligne]. <http://www.cirad.fr/ur/systemes_canniens> (consultation le 20-07-2011)

FranceAgriMer, Etablissement national des produits de l'agriculture et de la mer. *Communiqué de presse*, [En ligne]. <http://docs.google.com/viewer?a=v&q=cache:ppZ9vAUDCgYJ:www.franceagrimer.fr/Projet-02/02etablissement/02Etab_conseil_spe/02Erab_conseil_spe_PDF/c-sucre/Comm_press-ConseilSucre_24032011.pdf+prix+d%27exportation+du+sucre+revu+%C3%A0+la+baisse+?&hl=fr&pid=bl&srcid=ADGEESjpimxukvoXF-ivTvzJxp7iJ1NN-m2IU9IKqx7SdibvM04PbO4BO_op47oUlvjeU0c->

rjO9qoySKQixqK5A3hjzcsslhAJJiE25GmoUDUzk67aXAdBRzGsYi4Hit1l6ucIPk2K8&sig=AHIEtbSI5a9FMJaohky69Pahb8ZACOD_KA> (consultation le 26-07-2011)

En 24 heures. *Le ministère du Commerce annonce la baisse du prix du sucre*, [En ligne]. <<http://cote-d-ivoire.en24heures.com/le-ministere-du-commerce-annonce-la-baisse-du-prix-du-sucre>> (consultation le 26-07-2011)

Sucrerie de bois rouge, Sucrière de La Réunion. *Campagne sucrière 2008*, [En ligne]. <http://canasuc.gqf.com/sr_fichiers/documents/campagne_2008_dossier_presse_final.pdf> (consultation le 27-07-2011)

Sucrerie de bois rouge, Sucrière de La Réunion. *Campagne sucrière 2008*, [En ligne]. <http://canasuc.gqf.com/sr_fichiers/documents/Bilan%20campagne%202009%20der.pdf> (consultation le 27-07-2011)

Insee, L'Insee et la statistique publique. *Evolution de la population (TER)*, [En ligne]. <http://www.insee.fr/fr/themes/document.asp?amp;reg_id=24&ref_id=13969> (consultation le 29-07-2011)

eRcane, Valoriser la ressource canne. *Le CERF devient eRcane*, [En ligne] <<http://www.ercane.re/>> (consultation le 04-08-2011)

eRcane, Valoriser la ressource canne. *Procédés industriels*, [En ligne] <http://www.ercane.re/index.php?option=com_content&view=article&id=41&Itemid=2&lang=fr> (consultation le 04-08-2011)

eRcane, Valoriser la ressource canne. *Sélection expérimentale*, [En ligne] <http://www.ercane.re/index.php?option=com_content&view=article&id=111&Itemid=9&lang=fr> (consultation le 06-08-2011)

eRcane, Valoriser la ressource canne. *Hybridation*, [En ligne] <http://www.ercane.re/index.php?option=com_content&view=article&id=138&Itemid=6&lang=fr> (consultation le 06-08-2011)

Comment ça marche.net. *Apache-Introduction*, [En ligne] <<http://www.commentcamarche.net/contents/apache/apacintro.php3>> (consultation le 12-08-2011)

Wikipédia. *Modèle entité-association*, [En ligne] <http://fr.wikipedia.org/wiki/Mod%C3%A8le_conceptuel_de_donn%C3%A9es> (consultation le 16-08-2011)

Développez.com, Club des professionnels de l'informatique. *Quelle est la différence entre MCD et MLD*, [En ligne] <<http://www.developpez.net/forums/d365937/general-developpement/conception/modelisation/schema/difference-entre-mcd-mld/>> (consultation le 16-08-2011)

Gruau, Cyril. *Conception d'une base de données*, [En ligne] <<ftp://ftp-developpez.com/cyril-gruau/ConceptionBD.pdf>> (consultation le 16-08-2011)

Développez.com, Club des professionnels de l'informatique. *PHPMyAdmin : exporter le résultat d'une requête dans un tableur ?*, [En ligne] <<http://www.developpez.net/forums/d184763/bases-donnees/mysql/requetes/phpmyadmin-exporter-resultat-requete-tableur/>> (consultation le 18-08-2011)

TABLES DES SIGLES

GWH : GigaWatt-Heure

OCM : Organisation Communautaire des Marchés

DOM : Département d’Outre Mer

Cirad : Centre de coopération International en Recherche Agronomiques pour le Développement

EPIC : Etablissement Public à Caractère Industriel et Commercial

VCAT : Volontaire Civil à l’Aide Technique

UPR : Unité Propre de Recherche

SAU : Surface Agricole Utile

CERF : Centre d’Essai, de Recherche et de Formation

PAR : Radiations Photosynthétiques active

BDD : Base De Données

SGBD : Système de Gestion de Base de Données

SGBDR : Système de Gestion de Base de Données Relationnelles

HTML : Hyper Text Mark-Up Language

HTTP : Hyper Text Transfer Protocol

PHP : Hypertext Preprocessor

MySQL : My Structured Query Language

GNU : Système d'exploitation libre

GPL : General Public License

MCD : Modèle Conceptuel de Données

MLD : Modèle Logique de Données

MPD : Modèle Physique de Données

ID : Identifiant

ANNEXES

Annexe 1 : Les reliefs de La Réunion



Annexe 2 : Limite des « Hauts » de La Réunion.



Annexe 3 : Schémas et photos de boutures

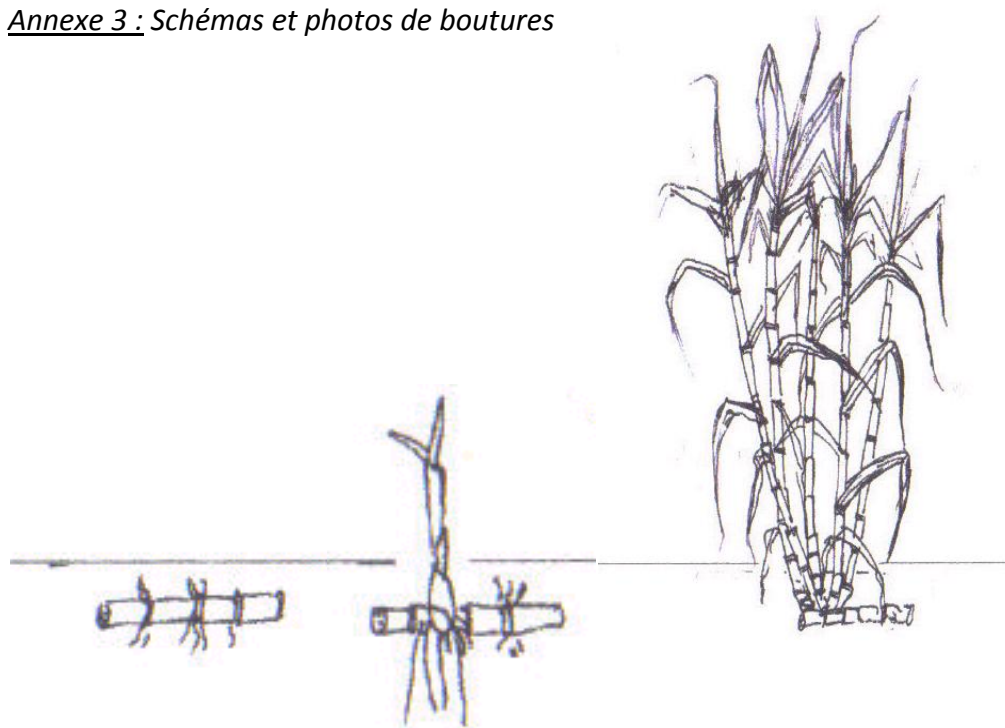


Schéma d'une bouture à 3 œilletons au stade plantation, 1 mois et 6 mois

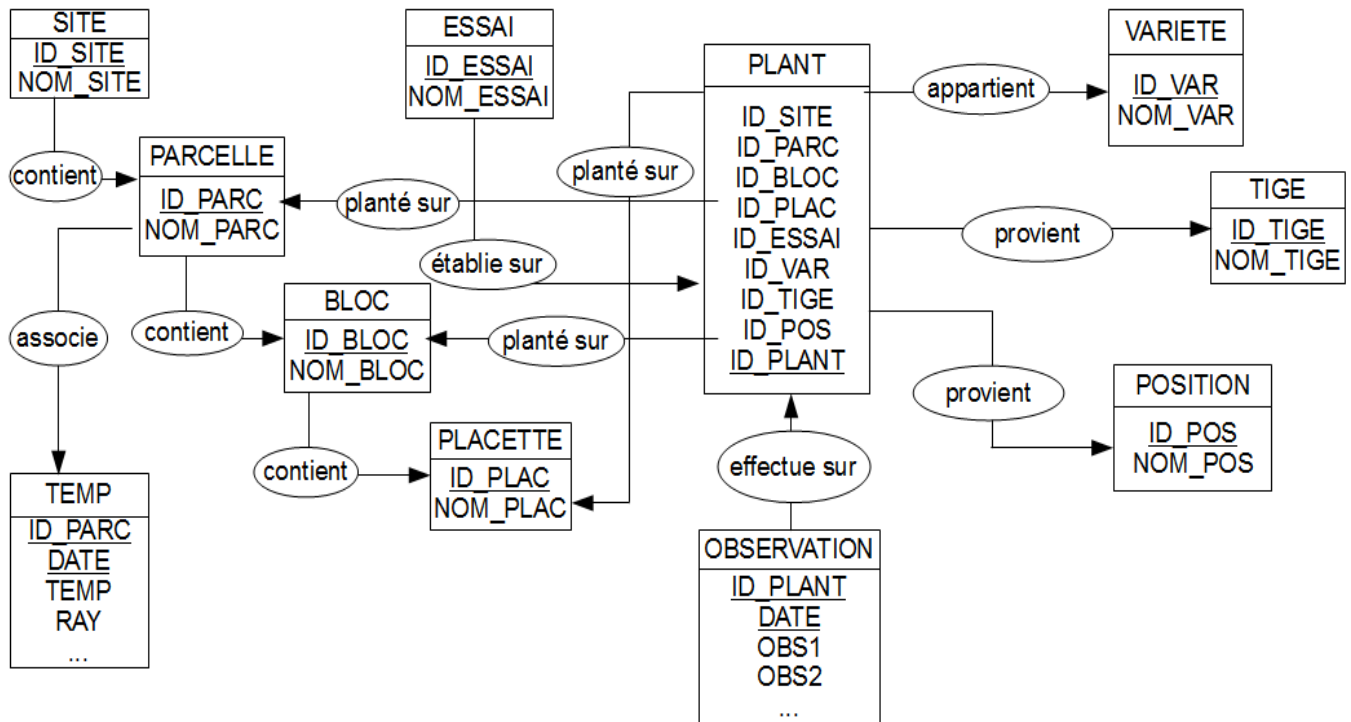


Photo d'une bouture à 1 œilleton



Photo de boutures à un œilleton, essai Germiléo

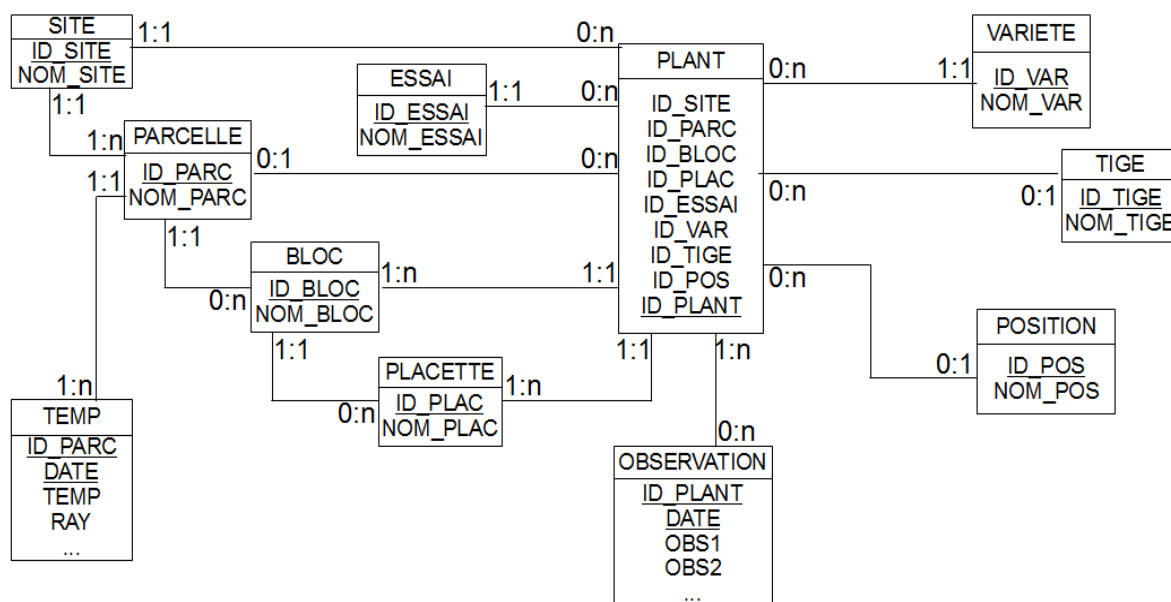
Annexe 4 : Le Modèle Conceptuel de Données



Annexe 4 :

MODELE CONCEPTUEL DE DONNEES DE LA BASE DE DONNEES

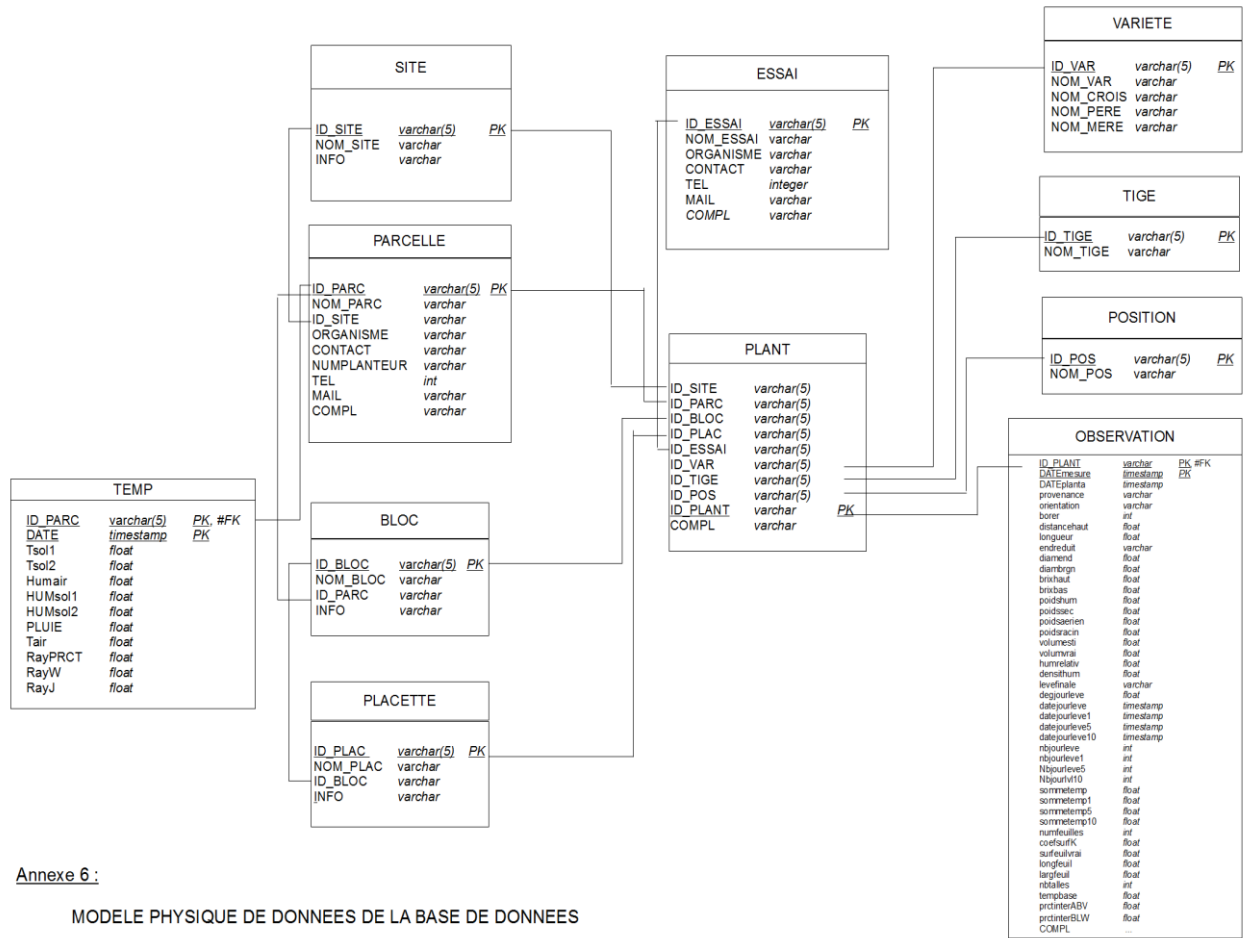
Annexe 5 : Le Modèle Logique de Données



Annexe 5 :

MODELE LOGIQUE DE DONNEES DE LA BASE DE DONNEES

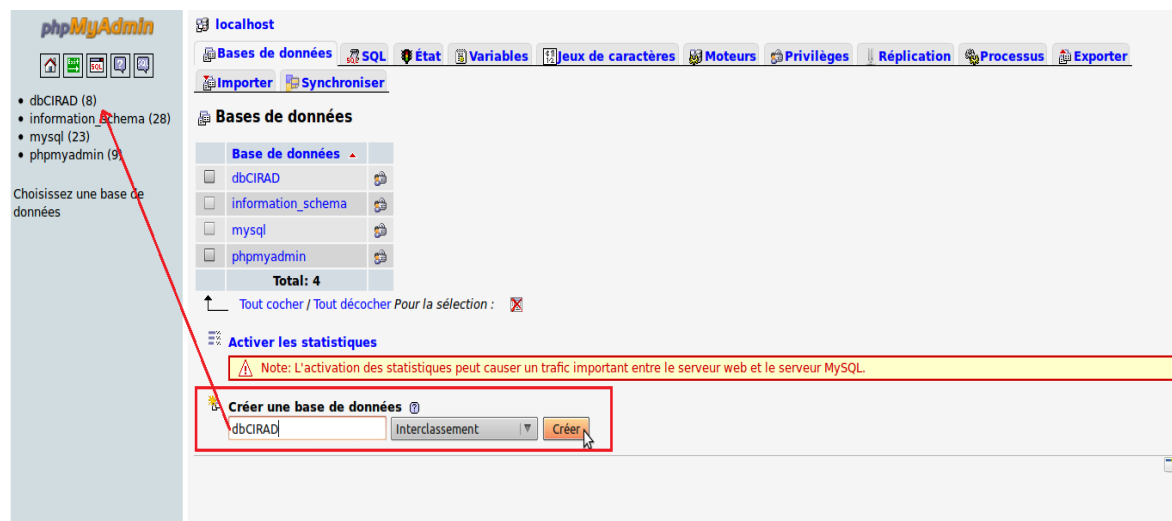
Annexe 6 : Le Modèle Physique de Données



Annexe 6 :

MODELE PHYSIQUE DE DONNEES DE LA BASE DE DONNEES

Annexe 7 : Création d'une nouvelle base avec PhpMyAdmin



Annexe 8 : Création d'une nouvelle table dans la base dbCIRAD avec PhpMyAdmin

